Matrix-free conditional simulations of Gaussian lattice random fields

DEBASHIS MONDAL

Department of Statistics,

Oregon State University,

Corvallis

In collaborations with Somak Dutta and Chunixao Wang

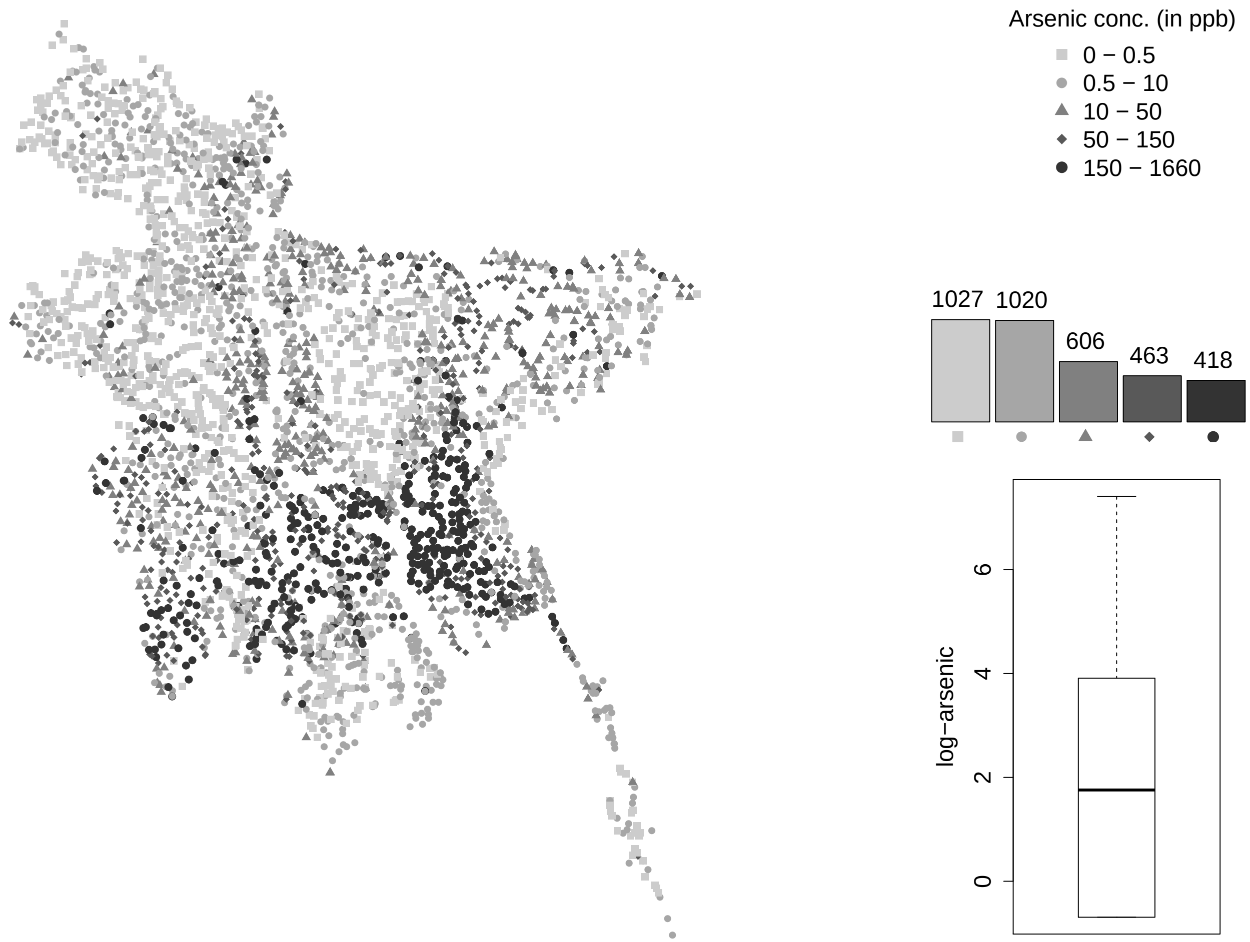
NSF CAREER Award number: 1519890

BIRS Workshop 19w5188

Outline for the talk

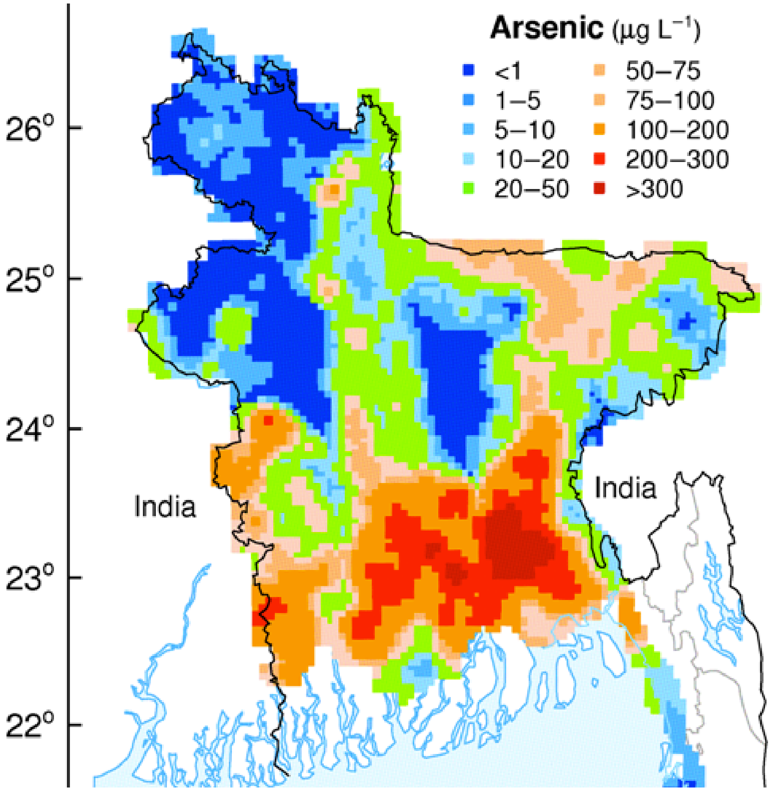
- An **overview** on prediction problems in **spatial statistics**
- Discuss **computational challenges**
- New algorithm for **matrix-free** for predictions on regular lattices
- **Extension** to spatial-temporal predictions
- Predictions on irregular lattices
- Applications in environmental sciences

Groundwater arsenic contamination in Bangladesh



Data at 3000 locations from a British Geological Survey

Groundwater arsenic contamination in Bangladesh

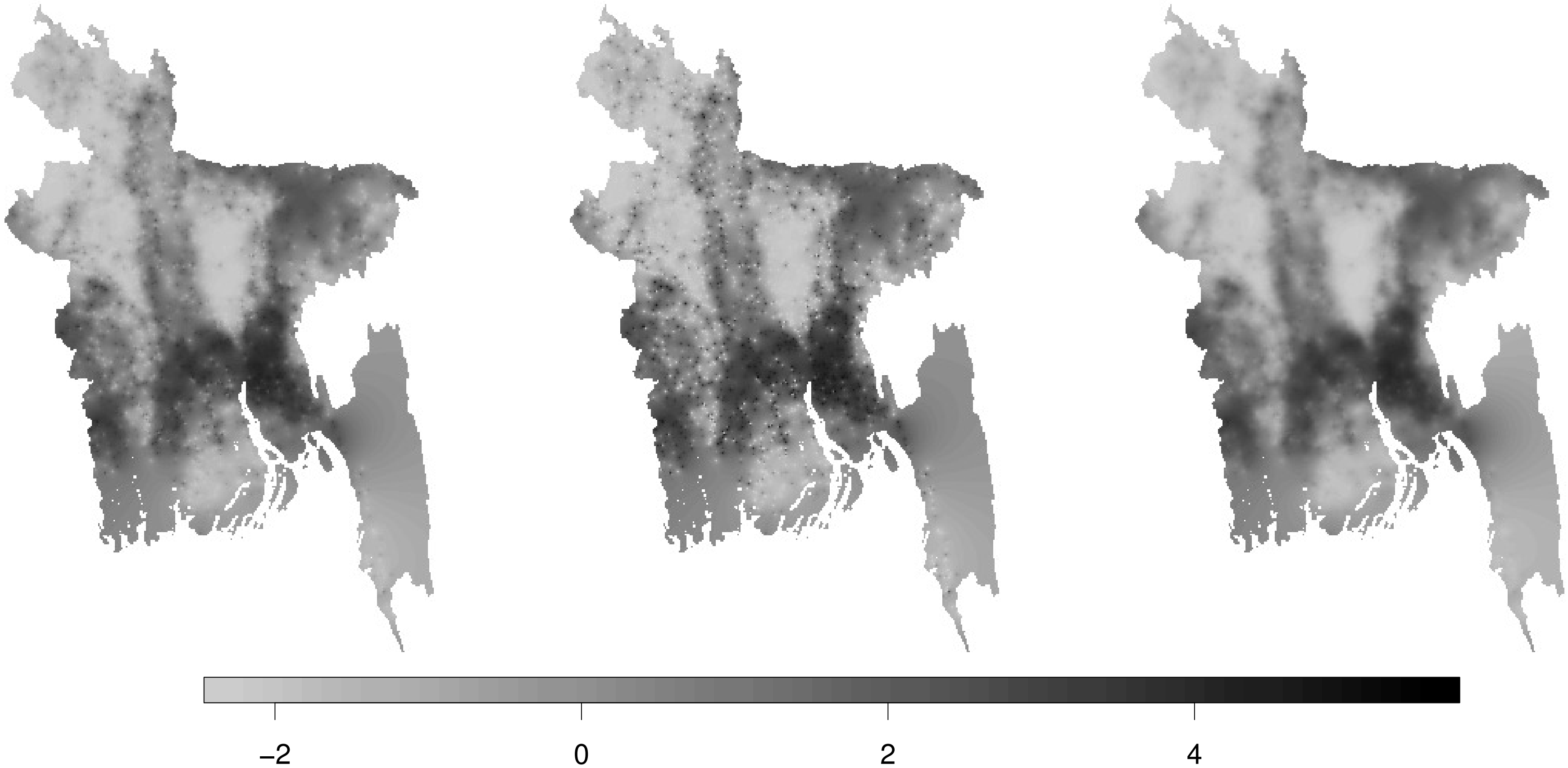


Arsenic contamination of the groundwater in **Bangladesh** is a serious problem.

Arsenic in **42 districts above WHO maximum permissible** limit of 50 mg/L.

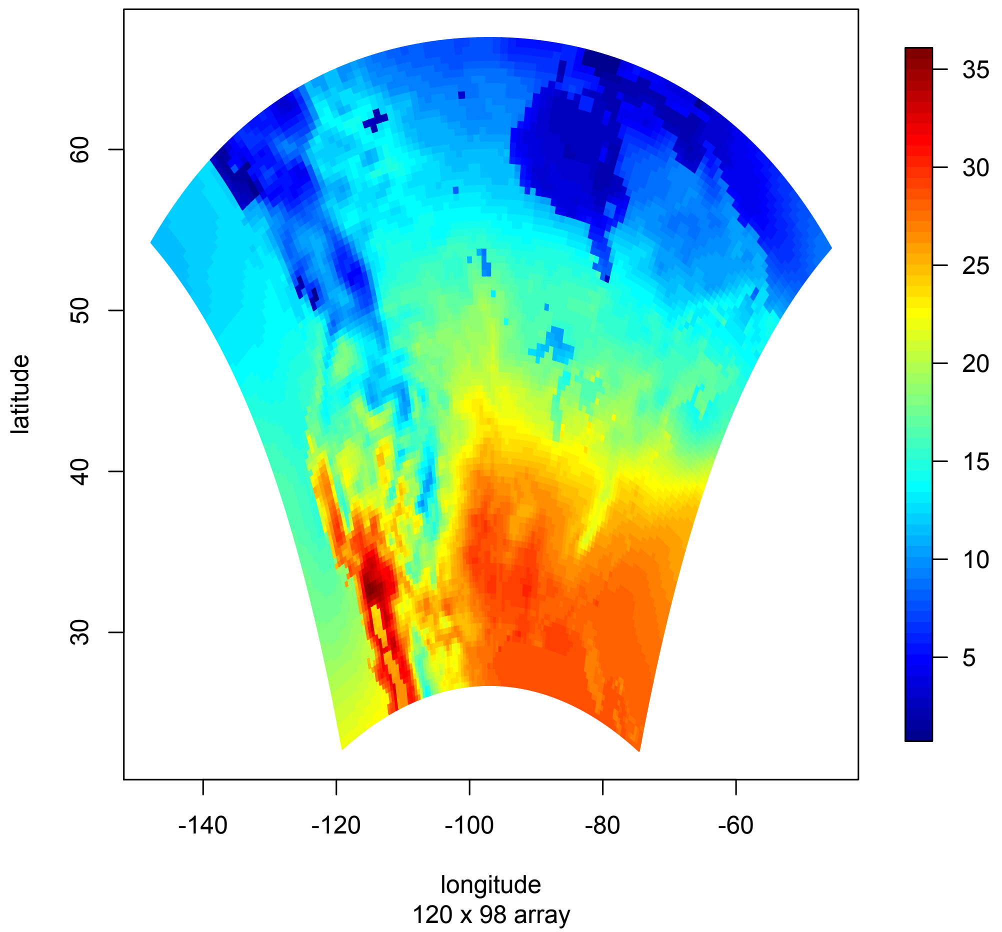
Many millions people are affected.

BLUPs for log arsenic contamination in Bangladesh



For **prediction**, we embedded the data on a 500×300 **grid**. Images corresponds to 3 different models. BLUPS are **starting points**, we also care about **uncertainties**.

Climate downscaling   Kaufman and Sain (2010)



Temperature from a regional climate model. **Goal** is to **predict** on a **finer resolution**.

EPA/EMAP study region II and sampling locations in 1994

Goal is to study the **extent of environmental damage**.

Prediction with spatial linear mixed models

$$y = T\tau + Fx + \epsilon$$

That is, we assume **Gaussian response** and **identity link** function.

y = $n \times 1$ vector of response

T = $n \times m$ covariate information matrix

τ = $m \times 1$ vector **covariate effects**

x = latent **spatial effects** on a **very fine** $r \times c$ **grid**

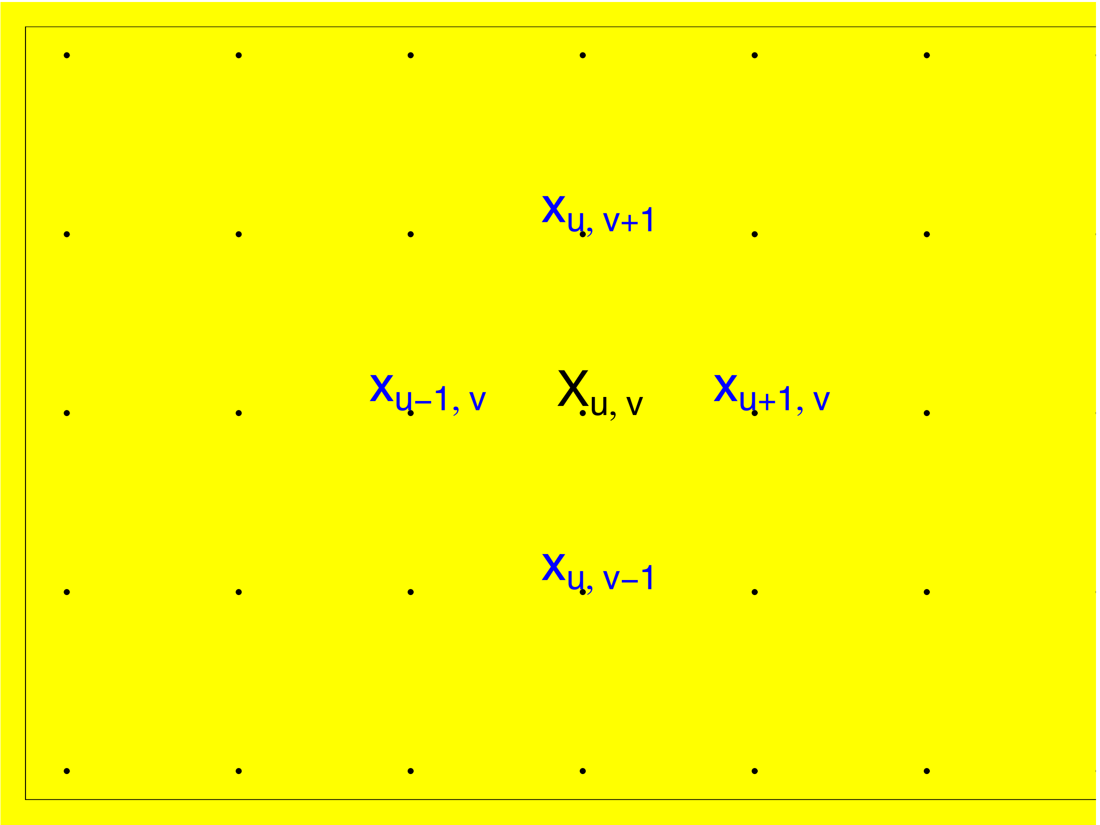
F = $n \times rc$ **sparse matrix** (identity or incidence matrix, averaging matrix,...)

ϵ = $n \times 1$ vector of Gaussian **residual** effects (might be omitted in some contexts)

No replication; data values not exchangeable; near values more related than distant ones

For prediction, we need stochastic modeling of x.

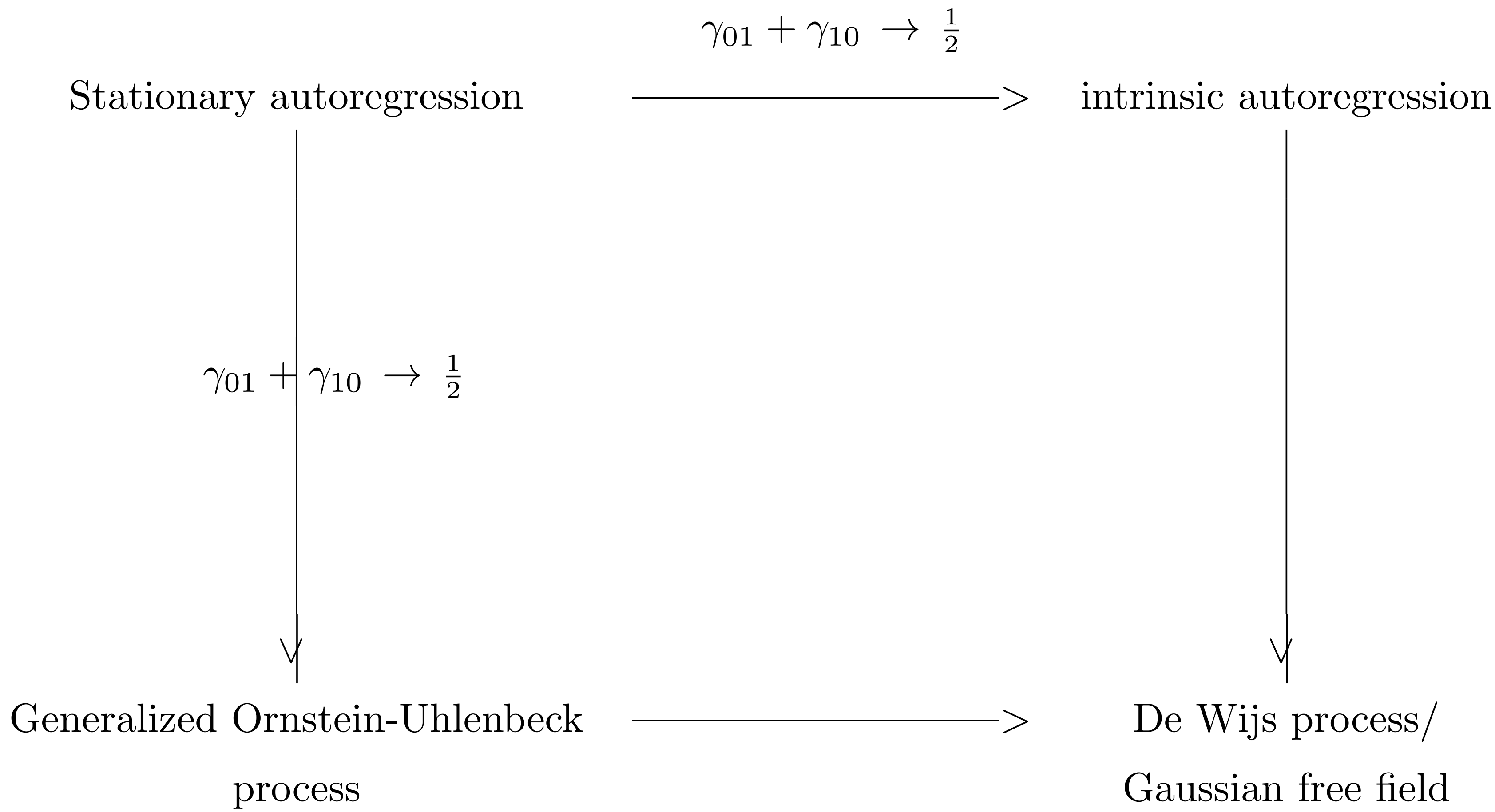
Nearest-neighbor (conditional autoregression) model for x



$$\mathrm{E}\left(X_{u,v} \mid \ldots\right) = \gamma_{10}\left(x_{u-1,v} + x_{u+1,v}\right) + \gamma_{01}\left(x_{u,v-1} + x_{u,v+1}\right), \qquad \mathrm{var}\left(X_{u,v} \mid \ldots\right) = \kappa.$$

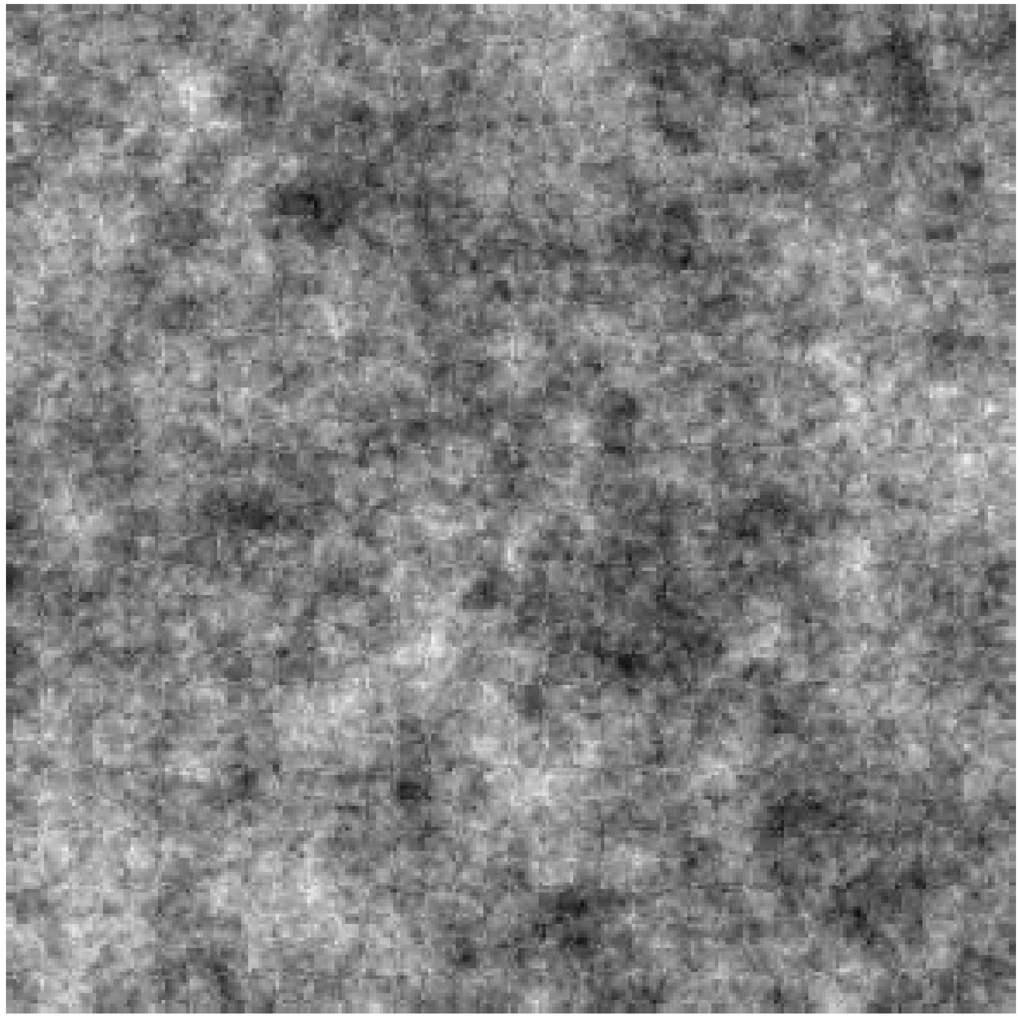
Typically $\gamma_{01}, \gamma_{10} \geq 0$ & $\gamma_{01} + \gamma_{10} \leq \frac{1}{2}$. **Focus on the intrinsic case** $\gamma_{01} + \gamma_{10} = \frac{1}{2}$.

Geostatistical limits of stationary and intrinsic autoregressions



- **A suitable** large **value of** s allows us to **approximate functionals of de Wijs process** by **functionals of Gaussian intrinsic autoregressions**

A realization of x on a 256×256 array



Further specifications of the linear mixed model

- Assume

$$\epsilon \sim \mathrm{N}\Big(0, \lambda_1^{-1} I_n\Big), \quad \tau \sim \mathrm{N}\Big(0, \lambda_2^{-1} I_m\Big) \tag{1}$$

- Distribution of x has an alternative form

$$|W|^{\frac{1}{2}} \exp\big\{-\tfrac{1}{2} x^T W x\big\}$$

  with two precision parameters λ_{10} and λ_{01} and

$$x^T W x = \lambda_{10} \sum \sum (x_{i,j} - x_{i-1,j})^2 + \lambda_{01} \sum \sum (x_{i,j} - x_{i,j-1})^2. \tag{2}$$

- **Precision matrix** W is **sparse** and has spectral decomposition is

$$W = MDM^T = M\Big(\lambda_{01} D_{01} + \lambda_{10} D_{10}\Big) M^T.$$

  M corresponds to the two dimensional **discrete cosine transformation**.

- Prior for dispersion parameters

$$\lambda \sim \pi(\lambda)$$

  Can consider **shrinkage** or other priors here

Conditional simulations

Interested in sampling from

$$\pi(\tau, x \mid y, \lambda) \equiv N\left(A^{-1}b, A^{-1}\right)$$

where

$$A = \begin{pmatrix} \lambda_1 T^T T + \lambda_2 I_m & \lambda_1 T^T F \\ \lambda_1 F^T T & \lambda_1 F^T F + W \end{pmatrix}, \quad b = \begin{pmatrix} \lambda_1 T^T y \\ \lambda_1 F^T y \end{pmatrix}$$

Current **state-of-the-art** algorithm

1.
$$z \sim N(0, I_{m+rc})$$

2. Compute sparse Cholesky decomposition
$$A = LL^T$$

3. Obtain by solving
$$(L^T)^{-1}L^{-1}b + (L^T)^{-1}z$$

Computational costs: **Memory** $= O((rc)^{\frac{1}{2}})$; **#FLOPs** $= O((rc)^{\frac{1}{2}})$. **Not scalable**!

A new algorithm

Consider a “rectangular square root”

$$A = SS^T, \quad S = \begin{pmatrix} \lambda_1^{\frac{1}{2}} T^T & \lambda_2^{\frac{1}{2}} I_m & 0 \\ \lambda_1^{\frac{1}{2}} F^T & 0 & MD^{\frac{1}{2}} \end{pmatrix}$$

1. Generate
$$z_1 \sim N(0, I_n), \quad z_2 \sim N(0, I_m), \quad z_1 \sim N(0, I_{rc})$$
2. Sample with A as covariance matrix, i.e. compute
$$u = Sz$$
3. Sample from $\pi(\tau, x \mid y, \lambda)$ by solving sparse equation
$$A\beta = b + u$$

Fast matrix-vector multiplication with A or S due to DCT

$A^{-1}b$ gives BLUE $\hat{\tau}$ and BLUP $\hat{x}$. Same **computational costs** as that of BLUP!

Lanczos algorithm + preconditioning with incomplete Cholesky

- To solve $A\beta = b$, A non negative definitive, use Lanczos algorithm

- sequentially compute orthonormal $v_1, v_2, v_3, \ldots$ from span of $b, Ab, A^2b, \ldots$ so that

$$AV \approx V\Delta$$

where Δ is tridiagonal, and obtain solution from a tridiagonal system of equation

- Matrix-free, depends only on matrix-vector multiplications

- Effective **order of computations** is $O(rc\log(rc))$.

- Preconditioning makes Lanczos algorithm even faster

- instead of solving $A\beta = b$ directly, solve:

$$CAC^T\beta' = Cb, \quad C^T\beta' = \beta$$

- One choice of C is block diagonal $\mathrm{diag}\{\lambda_1(T^TT)^{-\frac{1}{2}}, (\lambda_1F^TF + W)^{-\frac{1}{2}}\}$

- Replace $(\lambda_1F^TF + W)^{-\frac{1}{2}}$ by inverse **incomplete Cholesky** of $(\lambda_1F^TF + W)$.

Further developments

- For predictions we also need

$$\pi(\lambda \mid y) \approx N\Big(\lambda; \hat{\lambda}, \hat{I}(\hat{\lambda})^{-1}\Big)$$

Here $\hat{\lambda} \equiv$ **MLE** of λ, observed **Fisher information** $\equiv \hat{I}(\hat{\lambda})$,

See Dutta and Mondal (2015, 2016) for matrix-free computations of $\hat{\lambda}$ and $\hat{I}(\hat{\lambda})$

- Conditional simulations for **higher order intrinsic autoregressions**
  - E.g. thin plate splines: replace W by $P(W)$, P a positive polynomial.

Distribution of x has the form

$$|P(W)|^{\frac{1}{2}} \exp\big\{-\tfrac{1}{2} x^T P(W) x\big\}$$

$$A = SS^T = \begin{pmatrix} \lambda_1 T^T T + \lambda_2 I_m & \lambda_1 T^T F \\ \lambda_1 F^T T & \lambda_1 F^T F + P(W) \end{pmatrix}, \quad S = \begin{pmatrix} \lambda_1^{\frac{1}{2}} T^T & \lambda_2^{\frac{1}{2}} I_m & 0 \\ \lambda_1^{\frac{1}{2}} F^T & 0 & MP(D)^{\frac{1}{2}} \end{pmatrix}$$

Conditional simulations for intrinsic Matérn models

Typically, Matérn models defined via **covariances** involving **Bessel** functions ...

Focus here on a **discretized** version for which the distribution takes the form

$$|W^{\frac{1}{2}} \exp\left\{-\tfrac{1}{2}x^T W^\alpha x\right\}.$$

For this model

$$\pi(\tau, x \mid y, \lambda) \equiv N\left(A^{-1}b, A^{-1}\right)$$

where

$$A = SS^T = \begin{pmatrix} \lambda_1 T^T T + \lambda_2 I_m & \lambda_1 T^T F \\ \lambda_1 F^T T & \lambda_1 F^T F + W^\alpha \end{pmatrix}, \quad S = \begin{pmatrix} \lambda_1^{\frac{1}{2}} T^T & \lambda_2^{\frac{1}{2}} I_m & 0 \\ \lambda_1^{\frac{1}{2}} F^T & 0 & M D^{\alpha/2} \end{pmatrix}$$

W^α is **not sparse**, but **discrete cosine transformation** helps in computation

$$\pi(\alpha, \lambda) = ???$$

See Dutta and Mondal (2016) for **MLE** calculations ...

Lattice systems and approximation to advection-diffusions

Consider

$$\partial x(t,z)/\partial t = -(1/2)\{\mathscr{A} x_t\} + \delta(t,z),$$

$\delta(t,z)$ Gaussian, temporally uncorrelated, and

$$\mathscr{A} x(t,z) = 2\mu^T \partial x(t,z)/\partial z - tr\{\partial^2 x(t,z)/(\partial z \partial z^T)\}\Sigma + 2\tau x(t,z),$$

RHS: 1st term for **transportation**, 2nd term for **diffusion**, 3rd term for **dumping**

For brevity, take $\mu = 0$, $\Sigma = \gamma_1 I$, $\tau = -\gamma_2$. Then

$$\partial x(t,z)/\partial t = \gamma_1\{\partial^2 x(t,z)/\partial z_1^2 + \partial^2 x(t,z)/\partial z_2^2\}/2 - \gamma_2 x(t,z) + \delta(t,z),$$

Discretization with $x_{i,z} = x(i\Delta, z\Delta_0)$ gives

$$\begin{aligned}&\Delta^{-1}(x_{i+1,z} - x_{i,z})\\ &= \gamma_1 \Delta_0^{-2}(x_{i,z_1+1,z_2} + x_{i,z_1-1,z_2} + x_{i,z_1,z_2+1} + x_{i,z_1,z_2-1} - 4x_{i,z})/2 - \gamma_2 x_{i,z} + \delta_{i,z}.\end{aligned}$$

So

$$x_{i+1} = Kx_i + \delta_i, \quad K = (1-\gamma_2\Delta)I_n - \gamma_1 \Delta {\Delta_0}^{-2}(I_c \otimes S_1 + S_2 \otimes I_r)/2 = f(W).$$

Inference and predictions for state-space models

$$y_i = F_i x_i + \epsilon_i, \quad x_i = K x_{i-1} + \delta_i, \quad i = 1, \ldots, t$$

$$\epsilon_i \sim N(0, \gamma_4^{-1} I), \quad \delta_i \sim N(0, \gamma_3^{-1} I), \quad K = f(W),$$

Typically done via **Kalman filtering**. However for large data requires either

- dimension reduction, or
- ensembles of stochastic simulations
- data sketching

Are **scalable, matrix-free, statistically efficient** predictions possible???

Spectral property of the inverse-covariance matrix

$$x^T = (x_1^T, \ldots, x_t^T), \quad \Gamma^{-1} = \mathrm{var}\,(x)$$

Let M and M^T correspond to **two-dimensional DCT** and **inverse DCT**. Then

$$f(W) = M f(D) M^T,$$

where D is diagonal, known and

$$\Gamma = R\Omega R^T,$$

where

$$R = \begin{pmatrix} M & & \\ & \ddots & \\ & & M \end{pmatrix}, \Omega = \begin{pmatrix} \gamma_3 I & -\gamma_3 f(D) & 0 & 0 & \cdots \\ -\gamma_3 f(D) & \gamma_3[1+f(D)^2] & -\gamma_3 f(D) & 0 & \cdots \\ 0 & -\gamma_3 f(D) & \gamma_3[1+f(D)^2] & -\gamma_3 f(D) & \cdots \\ \cdots & 0 & -\gamma_3 f(D) & \gamma_3[1+f(D)^2] & -\gamma_3 f(D) \\ \cdots & 0 & 0 & -\gamma_3 f(D) & \gamma_3 I \end{pmatrix}$$

Can compute $R\theta$ or $\Gamma\theta$ in $O(rct\log(rct))$ steps with storing matrices!!

Conditional simulations of state vectors

Use vectorize forms

$$y^T = (y_1^T, \ldots, y_t^T), \quad \epsilon^T = (\epsilon_1^T, \ldots, \epsilon_t^T), \quad \zeta^T = (\zeta_1^T, \ldots, \zeta_t^T), \quad F = \mathrm{Diag}(F_1, \ldots, F_s)$$

$$\text{Then} \quad \pi(x \mid y, \gamma) \equiv N\left(A^{-1}b, A^{-1}\right), \quad A = \gamma_4 F^T F + \Gamma, \quad b = \gamma_4 F^T y$$

$$A = SS^T, \quad S = (\ \gamma_4^{\frac{1}{2}} F^T \quad RB\ ), \quad \Omega = BB^T, \quad B \text{ lower lower block bidiagonal}$$

$$B = \gamma_3^{\frac{1}{2}} \begin{pmatrix} I & 0 & 0 & 0 & \cdots \\ -f(D) & I & 0 & 0 & \cdots \\ 0 & -f(D) & I & 0 & \cdots \\ \cdots & 0 & -f(D) & I & 0 \\ \cdots & 0 & 0 & -f(D) & I \end{pmatrix}$$

Again computational cost is $O(rct\log(rct))$ **without storing matrices**!!

See Mondal and Wang (2019) for **MLE** computation for γ.

Conditional simulations for spatial models on irregular lattices

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$ a d**ependence** graph; $i \sim j \Leftrightarrow i$ neighbor of j; $\partial i \Leftrightarrow$ all neighbors of i.

$\omega_{i,j}$ **proximity** measure between i and j; $\omega_{i,j} > 0$ if $i \sim j$; $\omega_{i,j} = 0$ if $j \notin \partial i$.

Consider spatial model x such that

$$\mathrm{E}\left(x_i \mid_{-i}, \lambda_0\right) = \sum_{j \in \partial i} \frac{\omega_{i,j}}{\omega_{i+\ j}}, \quad \mathrm{var}\left(x_i \mid_{-i}, \lambda_0\right) = \frac{1}{\lambda_0 \omega_{i+}}$$

Then

$$W_{i,i} = \lambda_0 \omega_{i,+}, \quad W_{i,j} = -\lambda_0 \omega_{i,j}, \quad \pi(x \mid \lambda_0) \propto \lambda_0^{n/2} |W|^{\frac{1}{2}} \exp\{-\tfrac{1}{2} \lambda x^T W x\}$$

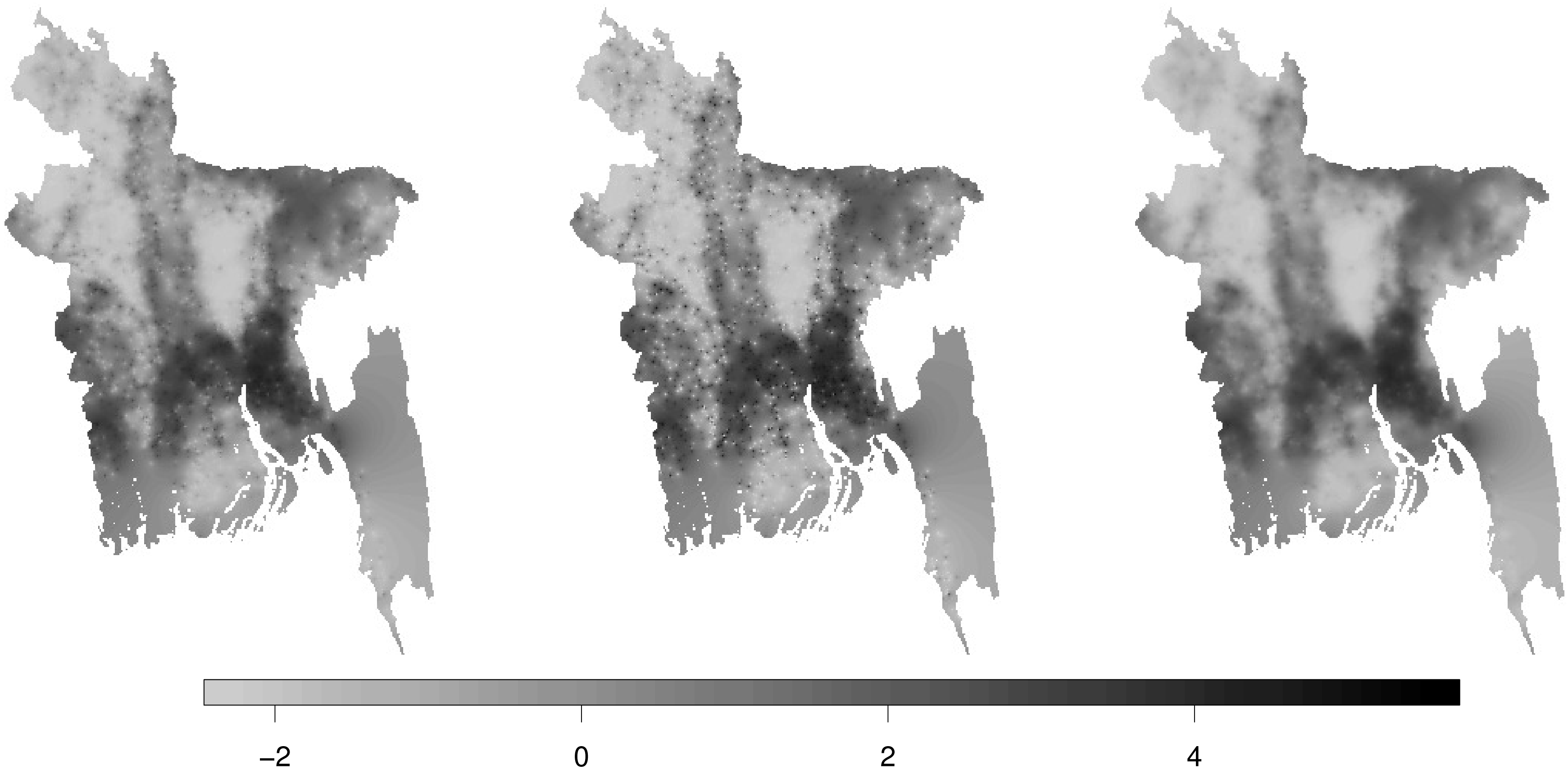
and

$$x^T W x = \sum_{i \sim j} \lambda_0 \omega_{i,j} (x_i - x_j)^2 = x^T \Big(\sum_{i \sim j} \lambda_0 \omega_{i,j} (e_i - e_j)(e_i - e_j)^T \Big) x = x^T B B^T x$$

It follows that

$$\pi(\tau, x \mid y, \lambda) \equiv N\left(A^{-1} b, A^{-1}\right), \quad \text{and } A = S S^T, \quad S = \begin{pmatrix} \lambda_1^{\frac{1}{2}} T^T & \lambda_2^{\frac{1}{2}} I_m & 0 \\ \lambda_1^{\frac{1}{2}} F^T & 0 & B \end{pmatrix}$$

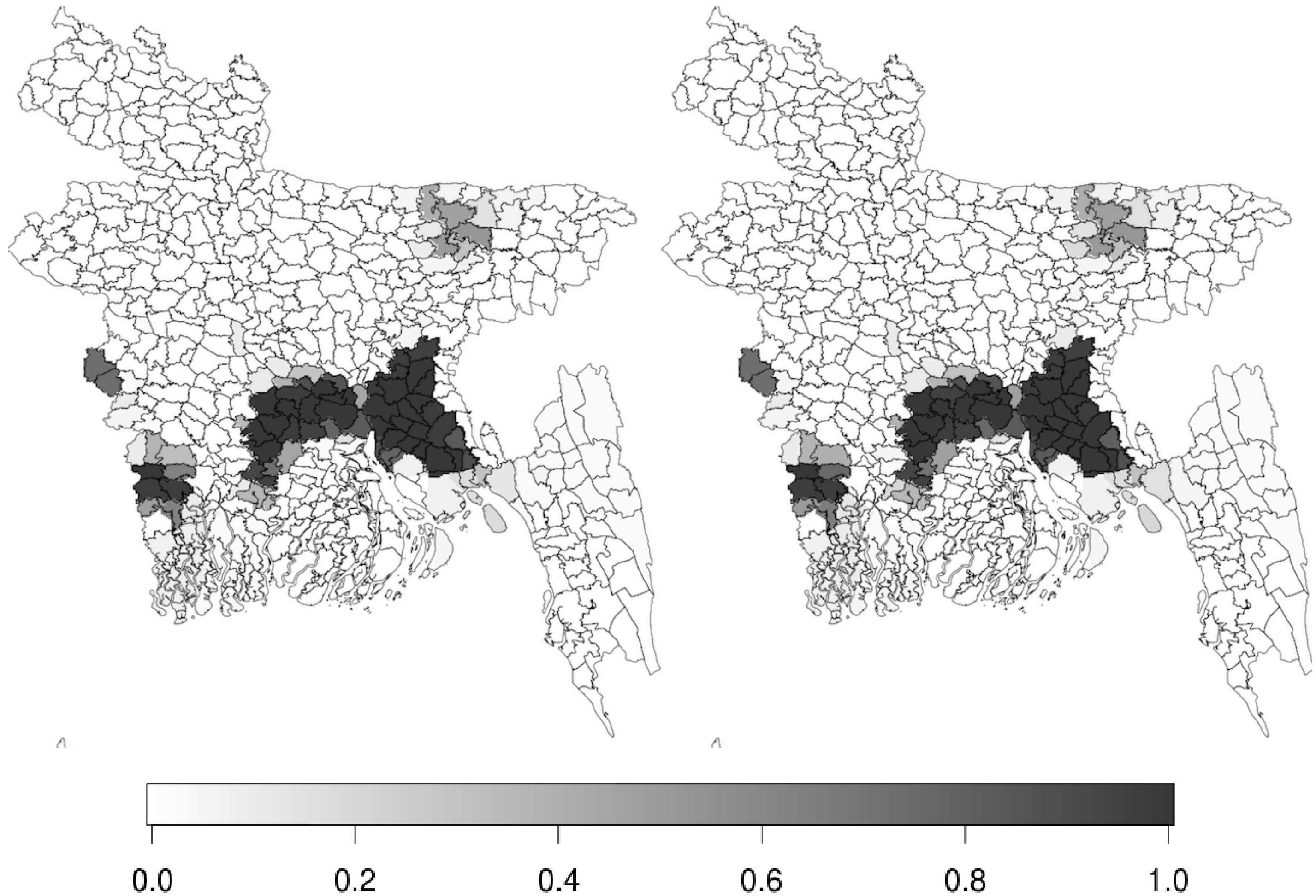
Back to log arsenic contamination in Bangladesh



About 3000 observations. BLUPs on a 500×300 grid.

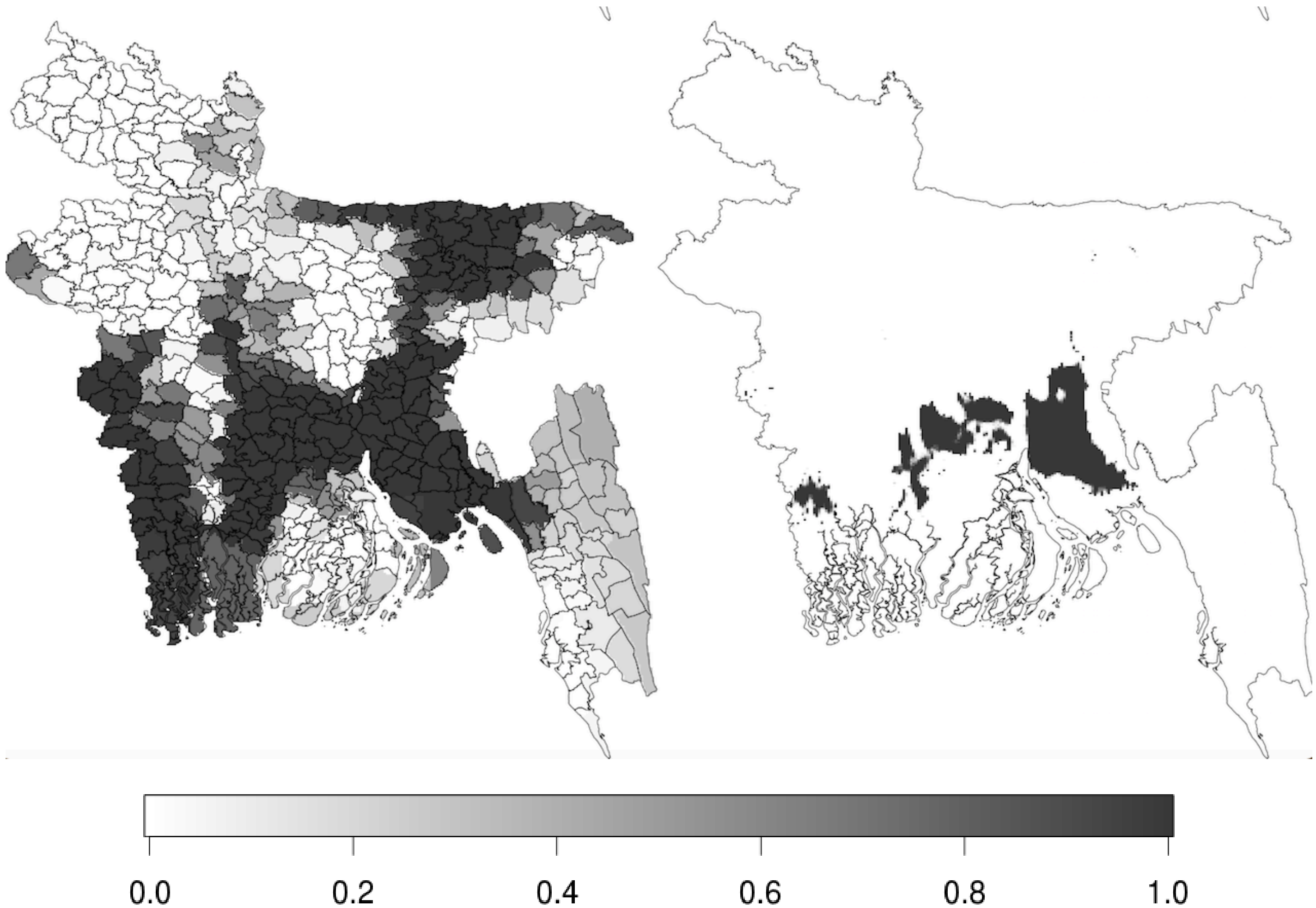
Left: for $\alpha = 1$, Center: for $\widehat{\alpha} = 0.858$ (no nugget), Right: for REML est. $\widehat{\alpha} = 1.240$

Pr(aggregated arsenic concentration exceeding 50ppb | data)



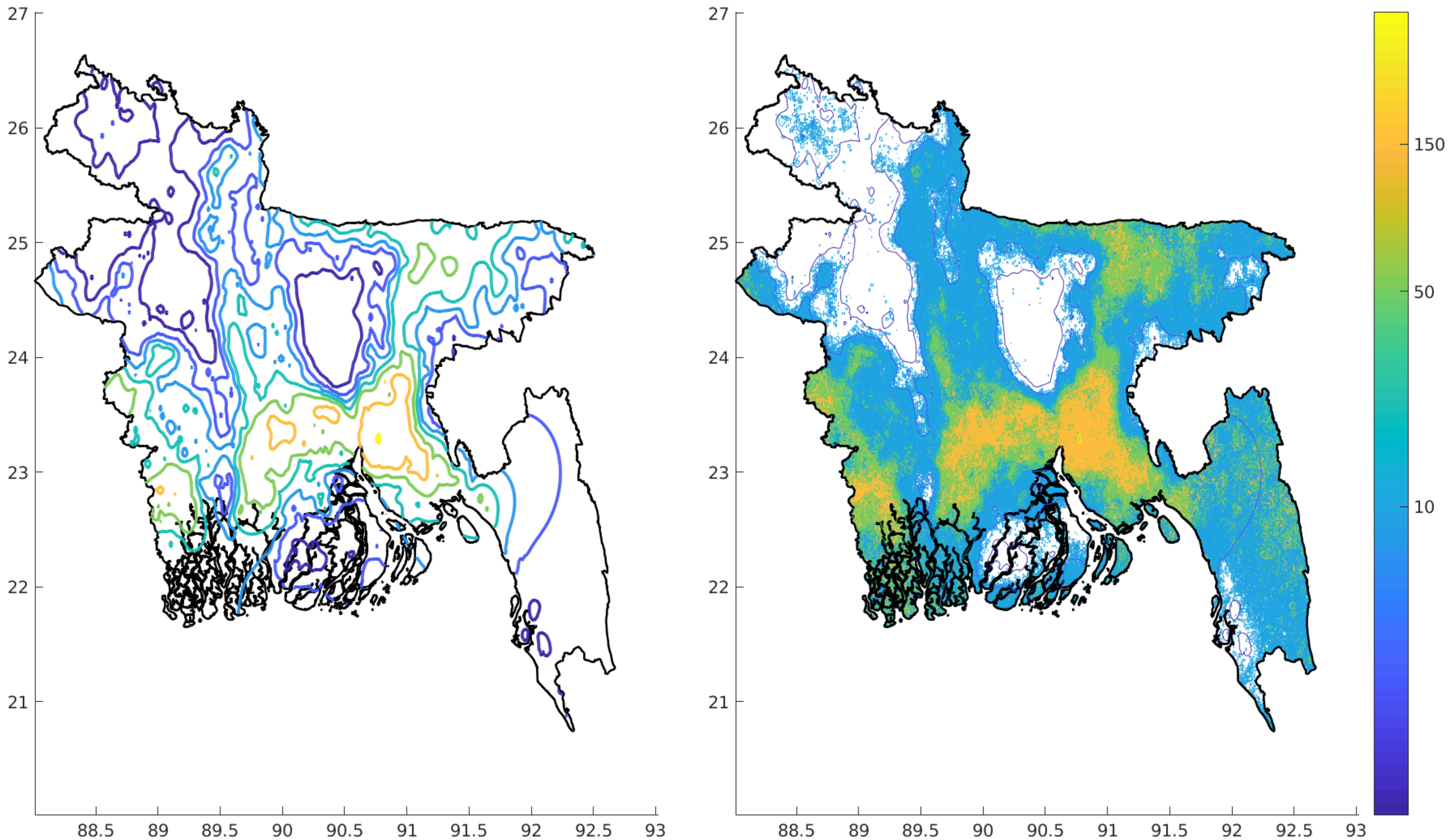
Geometric mean (left), median (right)

Pr(aggregated arsenic concentration exceeding 50ppb | data)



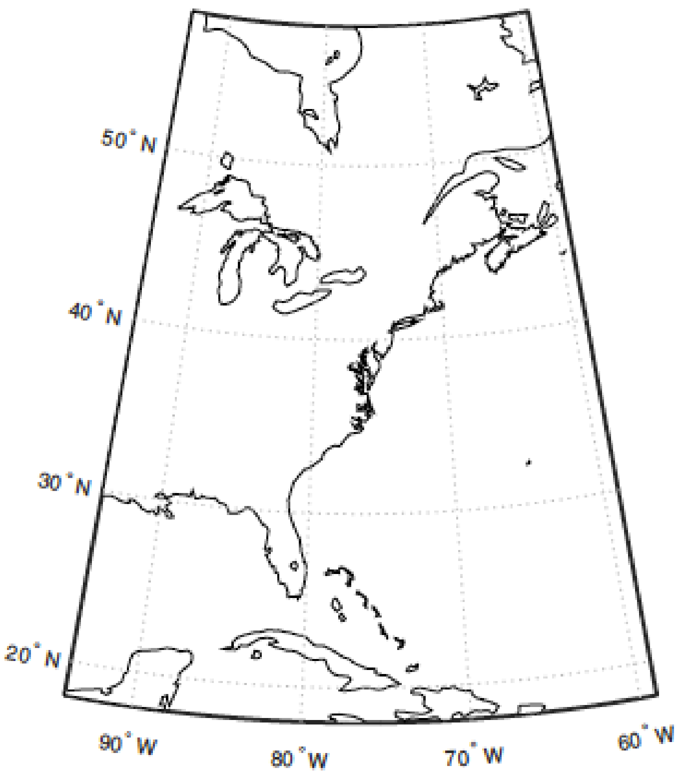
Maximum (left), inclusion probabilities in the maximal exceedance region (right)

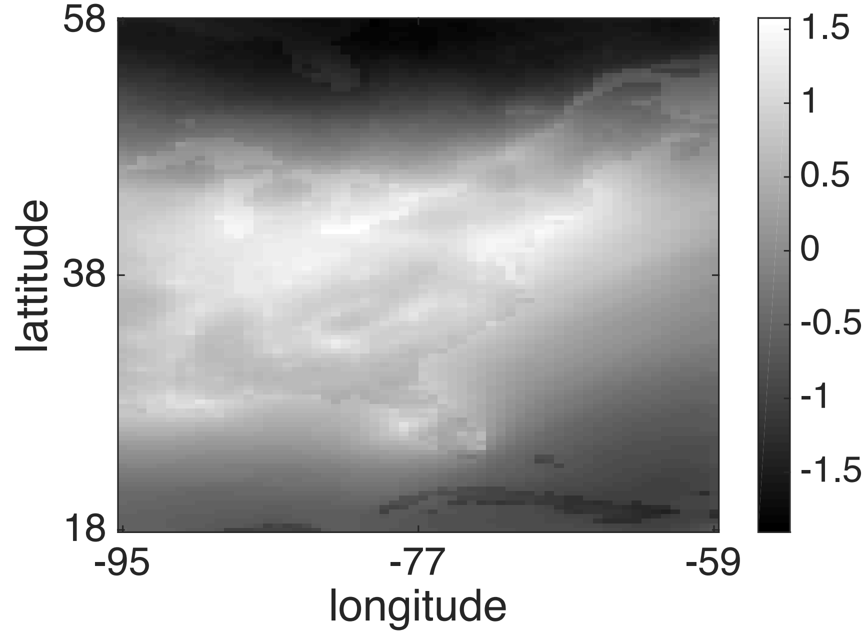
Visualization with contour lines etc.



Purple line: 0.5ppb; blue line: 10ppb, green line: 50ppb and yellow line: 150ppb

Standardized atmospheric concentrations of total nitrate

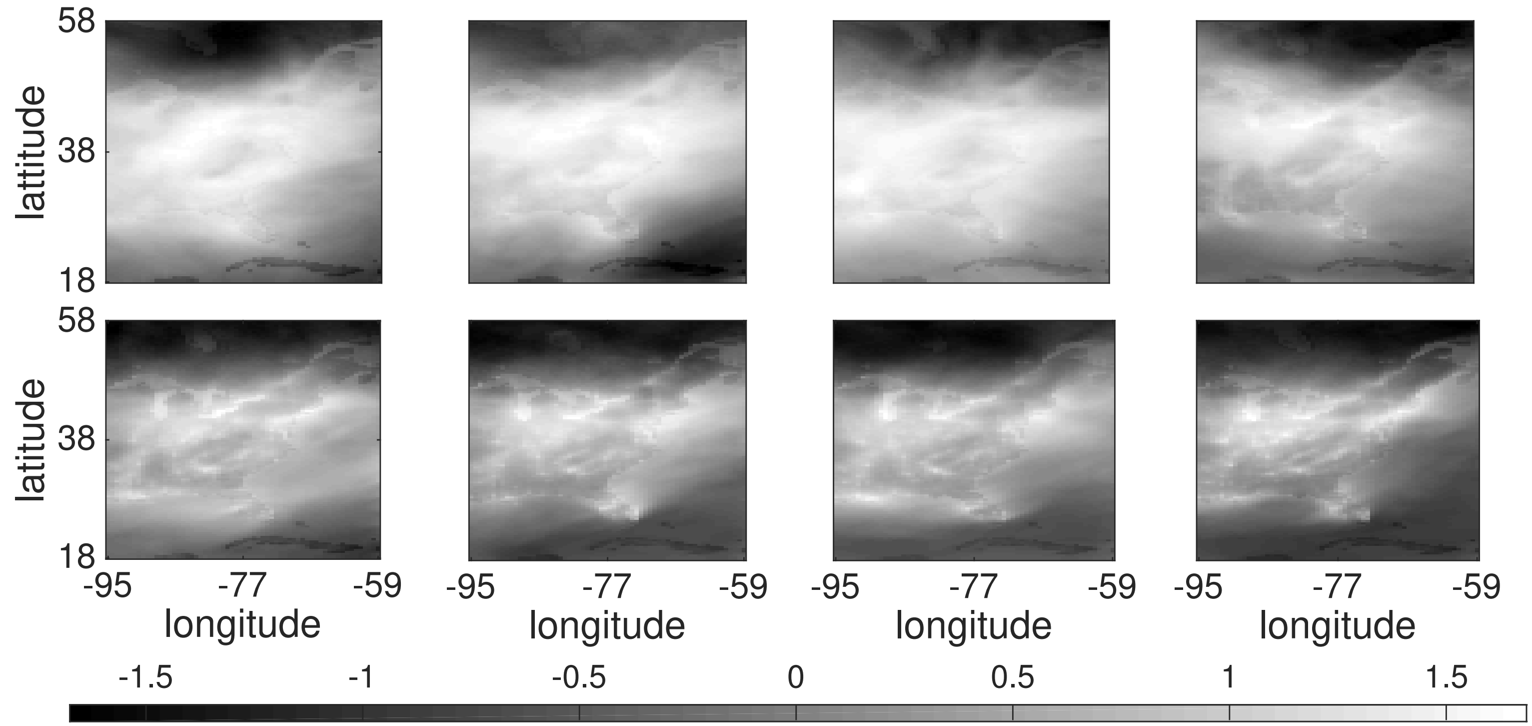




Models-3 output from EPA on gas-phase nitric acid plus particle-phase nitrate

Right panel gives average total Nitrogen concentration over 12 months in 2001.

Standardized atmospheric concentrations of total nitrate



Standardized Models-3 output for 8 lunar cycles

Spatial array size 62×112. Spatial resolution of each array cell 36×36 km^2.

REML estimates of precision parameters with standard errors

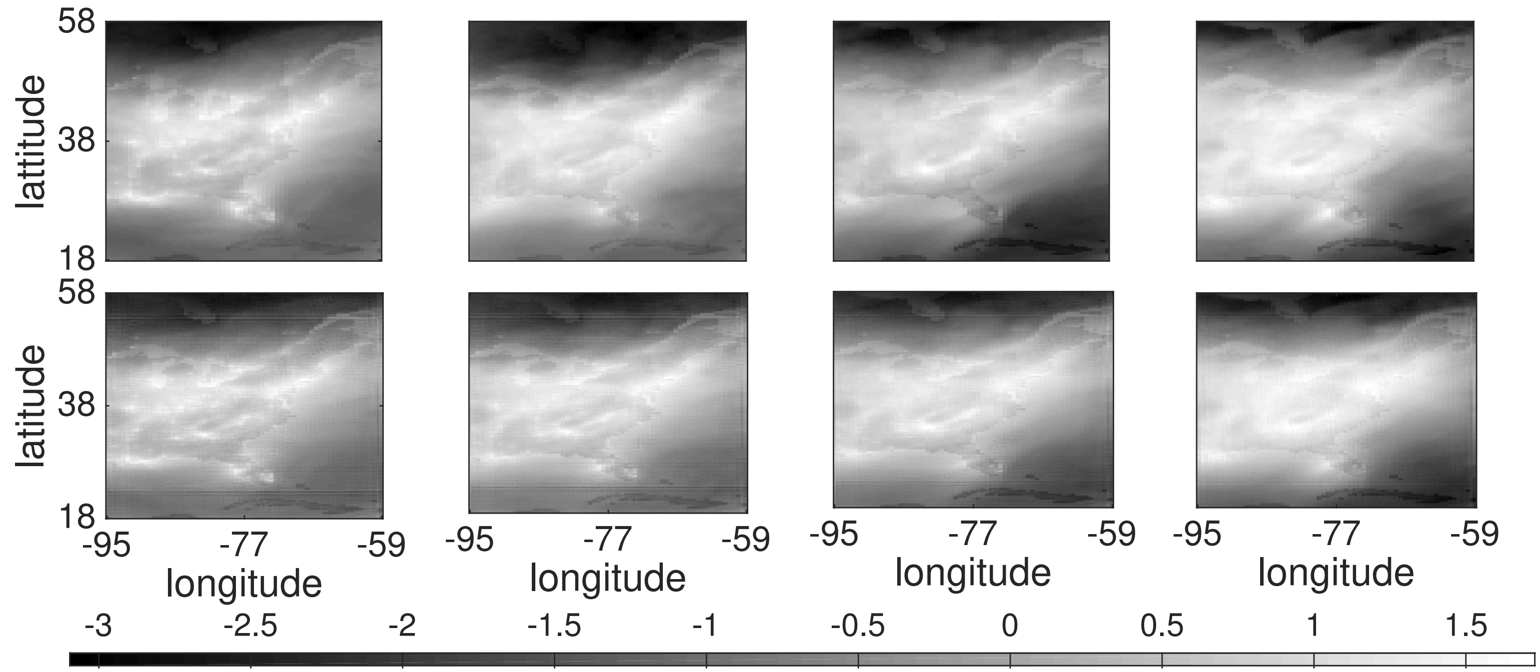
Parameters	λ_0	λ_1	λ_3
Scenario 1	7.536 (0.062)	7.894 (0.088)	14.571 (0.062)
Scenario 2	28.726 (0.118)	2.285 (0.028)	13.931 (0.119)

Based on fitting stochastic **advection-diffusion** equation

$$\mu = 0, \quad \Delta = 0.01, \quad \Delta_0 = 1, \quad \gamma_1 = \lambda_0\lambda_1/2, \quad \gamma_3 = 1/(\lambda_0\Delta), \quad \gamma_4 = \lambda_3, \quad \hat{\gamma}_2 = 0.$$

- Standard errors **in parenthesis**. Scenario 2 splits each pixel into 2×2 sub-pixels

Prediction of total nitrate at $18 \times 18\ km^2$ spatial resolutions



Top panel shows $y_9, \ldots, y_{12}$. Bottom panel displays $\hat{\psi}_9, \ldots, \hat{\psi}_{12}$.

Model explains about 96% of the total variations in the data.

REFERENCES

Anitescu, M. Chen, J. and Wang, L. (2012). A Matrix-Free Approach For Solving The Gaussian Process Maximum Likelihood Problem. To appear in *SIAM Journal of Scientific Computing.*

Besag, J. E. and Mondal, D. (2005). First–order intrinsic autoregressions and the de Wijs process. *Biometrika*, **92**, 909–920.

Besag, J. E. and Higdon, D. M. (1999). Bayesian analysis of agricultural field experiments (with discussion).
J. R. Statist. Soc. B, **61**, 691–746.

Besag, J. and Green, P. J. (1993). Spatial Statistics and Bayesian Computation (with discussions). *Journal of the Royal Statistical Society, B.* **55**, 25–37.

Borici, A. (2000). A Lanczos Approach to the Inverse Square Root of a Large and Sparse Matrix. *Journal of Computational Physics*, **162**, 123–131.

Dutta, S. and Mondal, D. (2016). REML estimation with intrinsic Matérn dependence in the spatial linear mixed model. Electronic Journal of Statistics, 10, 2856-2893.

Dutta, S. and Mondal, D. (2015). An h-likelihood method for spatial mixed linear models based on intrinsic autoregressions. Journal of Royal Statistical Society: Series B, 77, 699-726

Harville, D. A. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association.* **72**, 320–338.

Henderson, C. R. (1950). Estimation of genetic parameters. *Ann. Mlath. Stat.* **21**, 309–310.

Lee, Y. and Nelder, J. A. (2001). Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika*, **88**, 987–1006.

McCullagh, P. and Clifford, D. (2006) Evidence for conformal invariance of crop yields. *Proc Roy Soc A.* 2119-2143

Mondal, D. and Wang, C. (2017). Matrix-free computations of space-time Gaussian autoregressions and related processes. To appear in *Statistica Sinica*

Mondal, D. (2017). Generalized Gaussian Markov random fields and modeling disease risk. Under revision.

Paige, C. C., and Saunders, M.A. (1975). Solution of sparse indefinite systems of linear equations. *SIAM Journal on Numerical Analysis*, **12**, 617–629.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields. Theory and applications.* Chapman and Hall.